Digital Coding Of Audio Signals Using Perceptual Noise Criteria

by

R PARVATHAVARTHINI

DEPARTMENT OF ELECTRICAL ENGINEERING

Indian Institute of Technology Kanpur

JULY 1997

CERTIFICATE

This is to certify that the work contained in the thesis entitled Digital Coding Of Audio Signals Using Perceptual Noise Criteria by R Parvathavarthini has been carried out under my supervision and that this work has not been submitted elsewhere for a degree

Preeti Rao

Dr Preeti Rao
Assistant Professor
Department of Electrical Engineering
Indian Institute of Technology Kanpur

Abstract

Digital transmission, storage and reproduction of audio signals as effectively as possible needs adaptation to the perceptual characteristics of the human auditory system. This is best accomplished in the frequency domain. A Transform Coder based on the Discrete Fourier Transform is implemented. Perceptual parameters relevant to low bit rate coding are computed for a segment of sampled audio signal.

Acknowledgements

I wish to record my profound sense of gratitude and sincere thanks to my thesis supervisor Dr Preeti Rao for her invaluable guidance suggestions and cooperation throughout the work

I acknowledge with gratitude my Parental Department All India Radio & Doordarshan for giving me this golden opportunity

I am very thankful to my batchmates friends and Image Processing labmates for their constant encouragement in all aspects

July 1997 **R Parvathavarthini**

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1 1 Need Of Low Bit Rate Coders

Digital sound processing and storage adopted in audio are providing excellent sound quality However converting an audio signal to a 16 bit digital format with appropriate redundancy for error correction and with a minimum sampling rate requires extremely extended bandwidth for signal transmission and storage the latter coupled with huge mass storage necessities The large bandwidth results in problems for radio transmission in particular so there is considerable interest in avoiding any redundancy in the signal other than for error correction purposes To achieve sound transmission or reproduction that is not only very good but also efficient all equipment has to be adapted to the characteristics of the final receiver in this case the human auditory system Any part of the transmitted signal that is not recognized by the auditory system shows bad matching to the receivers and provides necessary redundancy Therefore digital audio coders are designed to reduce the redundancy and irrelevancy in a signal for the purpose of bit rate reduction That is digital waveform coders[4] seek to be in a region where digitized signals are perceptually relevant and statiscally nonredundant

Low bit rate coders if available with the requisite performance could be used in a number of places such as for remote broadcast lines studio links satellite transmission of high quality audio and the like For instance a channel with 15 KHz bandwidth that

is perceptually transparent would be very useful to broadcast media who could use it to provide a remote link free of the noise and frequency response problems normally associated with leased and/or equalized lines

Focus on wider bandwidth signals of 15 KHz or 20 KHz has centered primarily on bit rate reduction techniques The perceptual Transform Coder [3] described in this thesis uses a large encoding delay and has considerably greater complexity but provides transparent coding of a wide class of 15 KHz bandwidth audio signals at 4 bits per sample

1 2 Overview Of The Thesis

In this thesis we have attempted to gain an understanding of Perceptual Transform Coding concept by implementing a DFT based Transform Coder and computing some of the relevant perceptual parameters of a segment of sampled audio data The coding algorithm is based on Johnston s paper[3]

In the next chapter we review the ISO/MPEG Audio standard and chapter 3 reviews about the basic psychoacoustic property exploited in perceptual audio coding Chapter 4 describes the implementation of DFT based Transform Coder Chapter 5 discusses directions in Digital Audio Broadcasting(DAB)

Chapter 2

ISO/MPEG Audio Standard

2.1 MPEG Audio Compression Standard

Digital coded audio requiring low bit rates has led the ISO/IEC(International Organization for Standardization/International Electrotechnical Commission) standardization body to establish MPEG(Moving Pictures Experts Group) standards for audio and video coding. MPEG Audio standard specifies the coded representation of high quality audio for storage media and the method for decoding[8]. The compression techniques use psychoacoustic models for predicting the human auditory response to the noise that is introduced by the coding scheme. Using these models, the characteristics of the compression scheme can be changed dynamically in order to minimize the audibility of these noise impairments.

The MPEG Audio standard is intended for application to digital storage media such as Compact Disc(CD), Digital Audio Tape(DAT) and magnetic hard disc. The storage media may either be connected directly to the decoder or via other means such as communication lines. MPEG Audio standard is intended for sampling rates of 32 KHz, 44.1 KHz and 48 KHz at bit rates in a range of 32-192 Kbits per sec per mono and 64-384 Kbits per sec per stereo audio channel.

Depending on the application, three different layers of the coding system with increasing coder complexity and performance can be used.

Layer I This layer contains the basic mapping of the digital audio input into 32 sub

bands fixed segmentation to format the data into blocks a psychoacoustic model to determine the adaptive bit allocation and quantization using block companding and formatting This layer is most appropriate for consumer applications such as digital home recording on tapes that is for those applications for which very low data rates are not mandatory

Layer II This layer introduces further compression with respect to Layer I by redundancy and irrelevance removal on scale factors and uses more precise quantization This layer has numerous applications in both consumer and professional audio such as Digital Audio Broadcasting(DAB) recording telecommunication and multimedia

Layer III This layer introduces increased frequency resolution based on a hybrid filter bank It adds nonuniform quantization adaptive segmentation and entropy coding of the quantized values for better coding efficiency This layer is most appropriate for telecommunication applications in particular with narrowband ISDN and for professional audio having high weights on very low bit rates

2 2 Basic Structure Of ISO/MPEG/ Audio Encoder

The basic structure of a perceptual audio coding scheme[6] is shown in Fig 2 1 Fig 2 1(a) illustrates the basic structure of ISO/MPEG/Audio Encoder

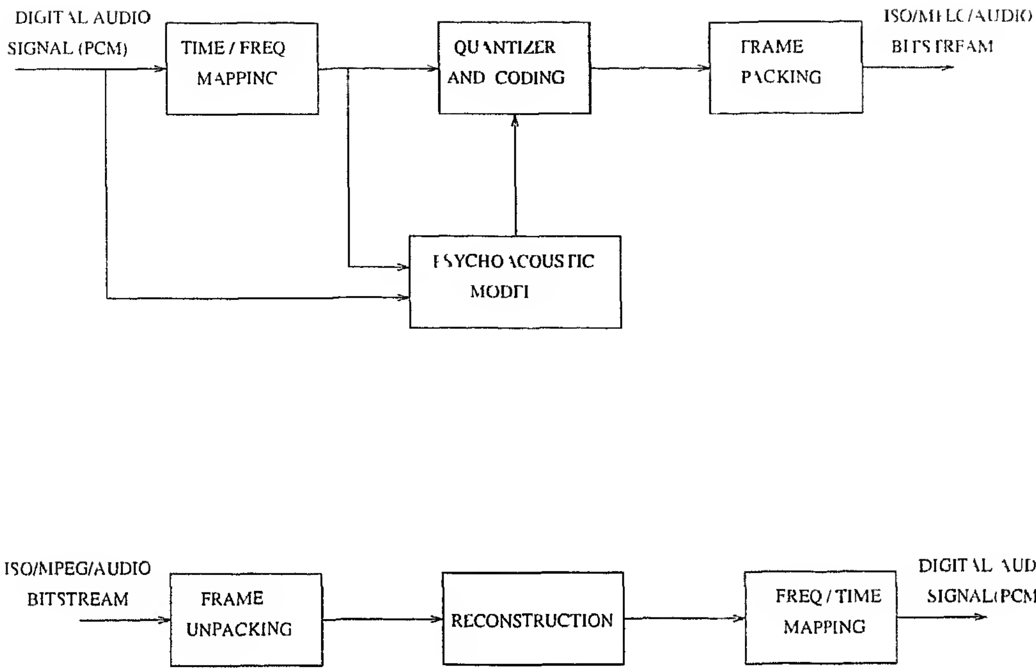


Figure 2.1 **(a) Audio Encoder (b) Audio Decoder**

It consists of

1. A time frequency mapping(filter bank) is used to decompose the input signal into subsampled spectral components. Depending on the filter bank used these are called subband values or frequency lines.

2. The output of this filter bank or the output of a parallel transform is used to calculate an estimate of the actual(time dependent) noise masking threshold using rules known from psychoacoustics.

3. The subband samples or frequency lines are quantized and coded with the aim of keeping the noise which is introduced by quantizing below the masking threshold.

4. A frame packing is used to assemble the bit stream which typically consists of the quantized and coded mapped samples and some side information such as bit allocation information.

This thesis focuses on high frequency resolution which leads to only limited time resolution Such systems are usually called Perceptual Transform Coders

2 3 Basic Structure Of ISO/MPEG/ Audio Decoder

Fig 2 1(b) illustrates the basic structure of ISO/MPEG/Audio Decoder Bit stream data is fed into the decoder The bit stream unpacking and decoding block does error detection if error check is applied in the encoder The bit stream are unpacked to recover the various pieces of information The reconstruction block reconstructs the quantized version of the set of mapped samples The frequency time mapping transforms these mapped samples back into uniform PCM

Chapter 3

Psychoacoustics And Perceptual Coding

3.1 Auditory Masking

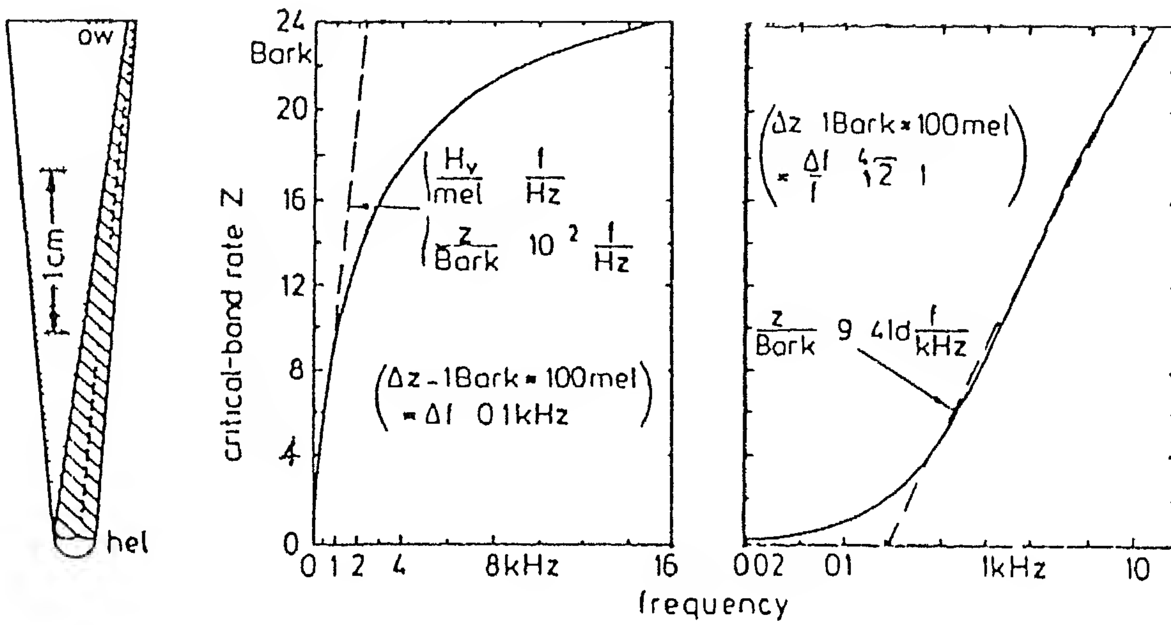
Most of the time our world presents us with a multitude of sounds simultanously. It is often difficult to hear one sound when a much louder sound is present. This process seems intutive but on the psychoacoustic and cognitive levels it becomes very complex. The term for this process is *masking*. Masking is defined as the amount by which the threshold of audibility for one sound is raised by the presence of another(masking) sound[10]. The basic idea in the psychoacoustic approach is to determine which part of the noise is masked by the signal and which part is audible.

Masking experiments using steady state signals have led to the conclusion that our auditory system performs a spectral analysis which can be modeled by a continuous filter bank of bandpass filters having a bandwidth of 100 Hz for frequencies below 500 Hz and a bandwidth of approximately one sixth of the center frequency[7] above 500 Hz. That is, our auditory system analyses in parts that correspond to critical bands. A list of critical band edge frequencies is given in Table 3.1.

Band No *Hz*	*Lower Edge* *Hz*	*Center* *Hz*	*Upper Edge* *Hz*
1	0	50	100
2	100	150	200
3	200	250	300
4	300	350	400
5	400	450	510
6	510	570	630
7	630	700	770
8	770	840	920
9	920	1000	1080
10	1080	1170	1270
11	1270	1370	1480
12	1480	1600	1720
13	1720	1850	2000
14	2000	2150	2320
15	2320	2500	2700
16	2700	2900	3150
17	3150	3400	3700
18	3700	4000	4400
19	4400	4800	5300
20	5300	5800	6400
21	6400	7000	7700
22	7700	8500	9500
23	9500	10500	12000
24	12000	13500	15500

Table 3 1 **A List Of Critical Band Center And Edge Frequencies**

Figure 3.1 (a) Scale of uncoiled cochlea (b),(c) Critical band rate Vs Frequency



From the Table 3.1 it is obvious that one critical band is added to the next so that the upper limit of the lower critical band corresponds to the lower limit of the next higher critical band thus producing the scale of the critical band rate. The critical band rate scale mentioned[1][5] follows a linear scale up to about 500 Hz and then a logarithmic frequency scale above 500 Hz.

Fig. 3.1(a) shows the uncoiled inner ear including the basilar membrane. It indicates that the critical band rate scale is directly related to the place along the basilar membrane where all the sensory cells are located in a very equidistant configuration. Thus the critical band rate scale is closely related to our physiology too. Fig. 3.1(b) and Fig. 3.1(c) shows the dependence of critical band rate on frequency as illustrated. The critical band concept has led to the definition of the *bark* scale as the psychoacoustic equivalent of the frequency scale.

3.2 Psychoacoustic Models

The MPEG audio standard is able to maintain CD audio quality reproduction up to a compression ratio of 5 to 1 or better. This bit rate corresponds to encoding each audio sample with an average of 3 to 4 bits per sample. The MPEG coding scheme achieves this compression by placing the quantization noise in those regions of the frequency spectrum where the human ear is least sensitive.

The psychoacoustic model is used to calculate the perceptual threshold for each time segment of the source material. This threshold, which estimates the maximum noise that can be inaudibly inserted into the signal, is used in the coder to maximize the amount of perceptually acceptable noise that can be inserted into a given signal segment. As the amount of the quantization noise is directly related to the number of bits used by the quantizer, the bit allocation algorithm assigns the available bits in a manner which minimizes the audible distortion.

Chapter 4

Implementation Of Perceptual Transform Coder

4 1 Why Transform Coder?

The Frequency Domain' approach which exploits linear dependencies between input samples for efficient digitization is called *Transform Coding* In this thesis Transform Coding is used to achieve the desired bit rate reduction for it is more robust than DPCM with regard to input statistics and channel errors The price paid for these advantages is increased encoding delay which is proportional to transform length N and an encoding complexity in the case with adaptive bit assignment

Fig 4 1(a) is a block diagram of transform coding It is a waveform digitizing procedure where a block of N input samples x[n] is linearly transformed into a set of N transform coefficients θ[n] The cofficients are then quantized for transmission and a reconstruction of x[n] is obtained at the receiver using an inverse transform operation on quantized coefficients

Fig 4 1(b) shows how the above coding scheme is implemented The input block is high resolution PCM samples The output of the coder is the combination of the outputs of 'N PCM coders that convey quantized uncorrelated coefficient information typically with a much lower total bit rate than what is present at the coder input

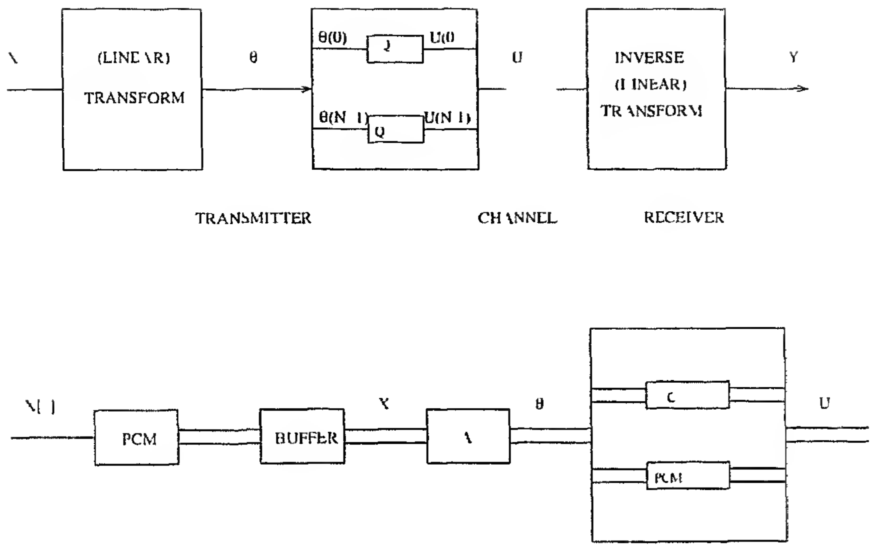


Figure 4.1 **(a)Transform Coder (b)Implementation of Encoder**

The efficiency of a transform coding system will depend on the type of linear transform and the nature of bit allocation for quantizing transform coefficients. Most practical systems will be based on Discrete Fourier Transform(DFT) approach.

DFT is defined by

$$\theta[k] = \sum_{n=0}^{N-1} x[n] \exp^{-j2\pi kn/N}, \qquad k = 0, 1, 2, \ldots, N-1 \qquad (4.1)$$

$$x[n] = 1/N \sum_{k=0}^{N-1} \theta[k] \exp^{-j2\pi kn/N} \qquad n = 0, 1, 2, \ldots, N-1 \qquad (4.2)$$

DFT representations have two important advantages

- DFT makes of sine cosine basis space. Hence it provides a natural framework for optimizations of transform coder design with inputs where the perceptual effects of signal distortion are best understood in the frequency domain.

- DFT is advantageous with respect to computation also A direct evaluation of an N point DFT requires about N^2 complex multiply add operations but if N is a power of 2 the Fast Fourier Transform(FFT) method needs in the order of $N \log_2 N$ operations With N=1024 the FFT is about one hundred times faster than a direct computation In the case of real valued inputs further reductions are possible by exploiting the properties of the DFT

- The FFT algorithm can also be used to evaluate the inverse operation

4 2 Implementation Of Transform Coder

A block diagram of the Transform Coder as implemented using *Matlab* package on Pentium PC @ 133 MHz, is shown in Fig 4 2 The source material is 16 bit PCM monophonic audio signal of 15 KHz bandwidth sampled at 32 KHz The window that is used in the overlap sections is the sqaure root of a Hanning window of length equal to twice the overlap that is 256 data points since the windowing process for overlap add and analysis provides for an overlap add of $\frac{1}{16}$th This overlap add window is shown in Fig 4 3 for three successive frames where each frame contains 2048 data points By allowing this overlap between the successive frames reduction of end effect noise is achieved without significantly lowering the number of bits available for encoding each frame With the overlap of $\frac{1}{16}$th the number of new data points in each frame is 1920

The FFT is implemented directly on the windowed data The 2048 real input data are converted to a spectrum of 1024 complex points counting the dc and Nyquist frequency as one complex point The spectral peaks of each 8 complex points are calculated Since the particular transform is a real complex transform of length 2048 there are 1024 unique complex pairs and 128 spectral peak values These spectral peak values are quantized

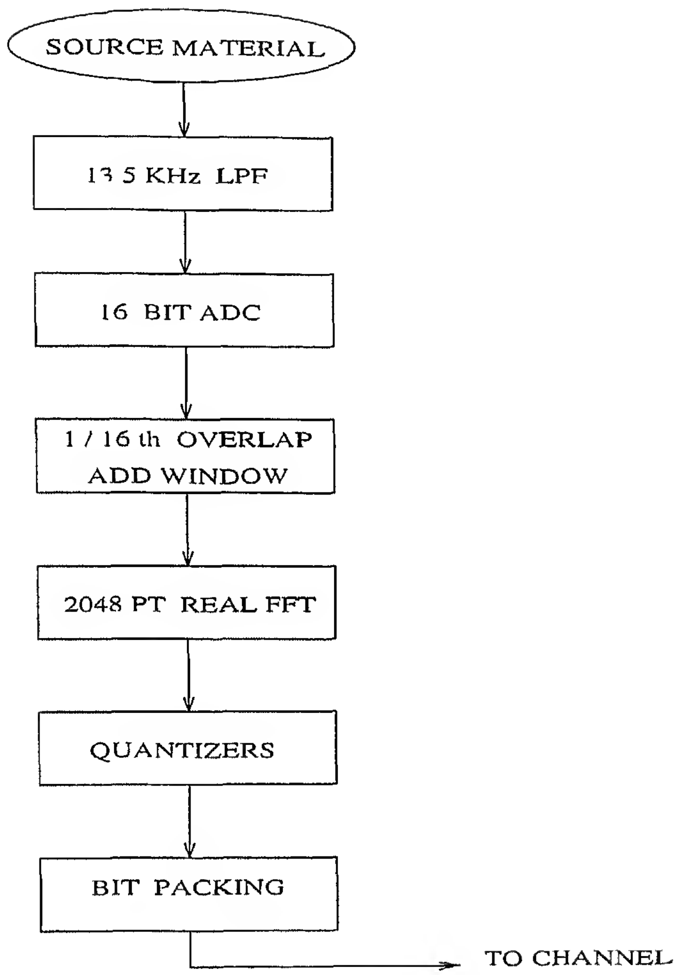


Figure 4 2 Block Diagram Of Transform Coder

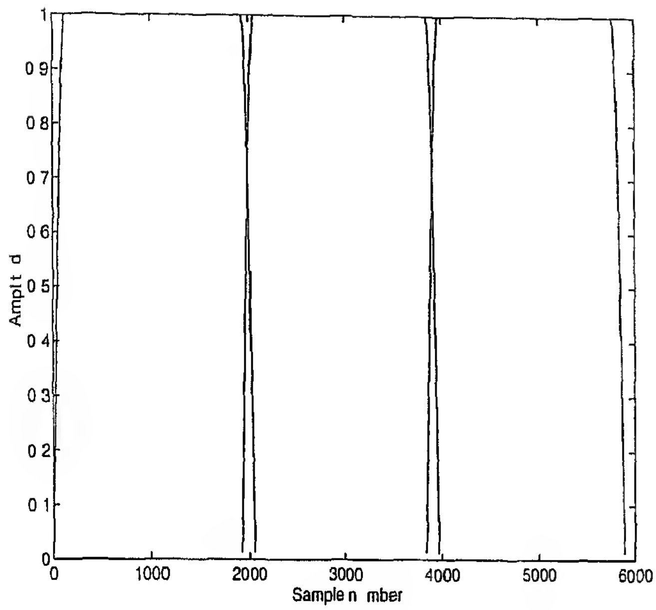


Figure 4.3 **Overlap add window**

using 8 bits, where each step size in the quantizer represents $\frac{170}{256} = 0.664dB$. The reason for choosing the number 170 is as follows.

The size of the quantizer is determined by the possible dynamic range of the 2048 point spectrum given a 16 bit PCM input source. In a transform, the total spectral dynamic range can vary from that of a unit impulse, here set to quantizer level 1, to the spectral energy of a full scale sinewave, roughly 167 dB higher, hence the total range of 170 dB.

Then the number of bits available for coding the samples are calculated. It requires the calculation of the following:

- The total number of bits available in a block of data.

- The number of side information bits required. The side information is used in both the transmitter and receiver to calculate bit allocation.

Since each block contains 1920 new data samples of 8 bits per sample the total number of bits available comes out to be 15360

The side information consists of three parts

- A 16 bit PCM encoding of the dc term from the FFT

- The threshold is encoded in 8 bit PCM where each step size is 0 664 dB

- The 128 spectral peaks are encoded in 8 bit PCM using step size of 0 664 dB

By taking into consideration the above details the number of bits required for side information is calculated to be 1040 This number is then subtracted from the total number of available bits to get the number of bits available for coding the samples

After calculating the actual number of bits available which is 14320 bits the number of levels in each set of quantizers is found by iterative procedure by calculating

$$k_i = 2 * nint(P_i/Thr_j) + 1 \quad (4\ 3)$$

where

k_i = number of levels in each set of quantizers

P = quantized peak level

Thr_j = quantized threshold value

nint = nearest integer function

j = Bark frequency, varying from 1 to 24

In this thesis threshold is initially assumed to be a constant and the number of levels in each set of quantizers is then calculated From the number of levels the number of bits is calculated and compared with the total number of bits available to encode a block of data samples If both numbers are not equal the threshold is adjusted accordingly and k is computed This process is repeated until all available bits have been used

The quantized samples along with the side information has to be transmitted to the receiver and reconstruction of the input signal x[n] is obtained at the receiver using inverse transform operation on quantized samples and the results windowed with the same window used for analysis Fig 4 4 shows the reconstructed signal of Transform Coder for 16 bit PCM input and Fig 4 5 shows the 8 bit PCM Coder output for the same input

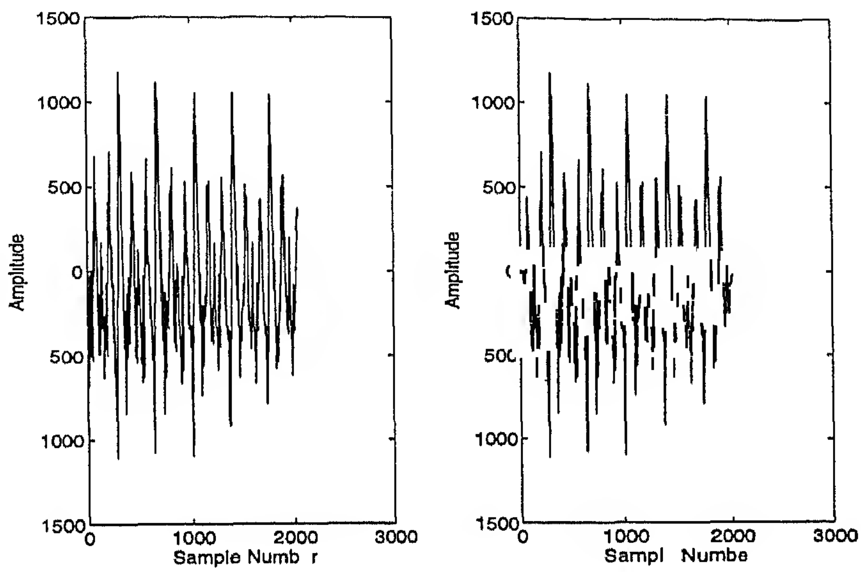


Figure 4 4 **(a)16 bit PCM input (b) Transform Coder output**

Listening to the reconstructed audio output reveals that while the 8 bit PCM coded segment is noticeably noisy the 8 bits per sample transform coded segment cannot be distinguished from the original

Also to compare their performance, Signal-to Noise ratio(SNR) is computed by first finding their respective quantization errors SNR computed in decibels for a block of input audio signal is as follows

$$SNR = 19\ 74\ dB \qquad For\ Transform\ Coder$$

$$SNR = 14\ 95\ dB \qquad For\ 8\ bit\ PCM\ coder$$

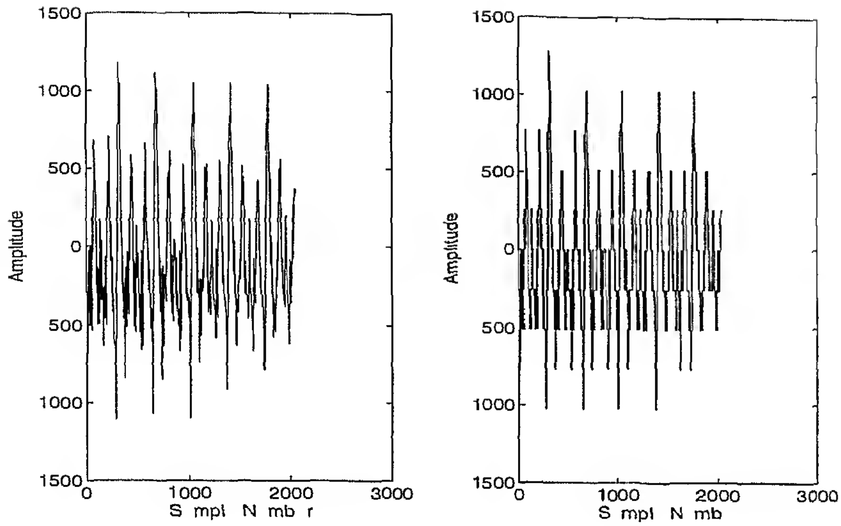


Figure 4 5 (a)16 bit PCM input (b) 8 bit PCM Coder output

The Threshold value used in Eqn 4 3 can be calculated using perceptual noise criteria Several steps are involved in calculating the masking threshold of which *Critical Band Analysis* using Bark spectrum and *Spectral Flatness Measure* are discussed

4 3 Critical Band Analysis

The block diagram of Perceptual Transform Coder is shown in Fig 4 6

From the 1024 unique complex pairs computed earlier (by windowing the input signal and processing by FFT), the power spectrum is calculated using

$$P(\omega) = Re^2(\omega) + Im^2(\omega) \tag{4 4}$$

The power spectrum is shown in Fig 4 9 The spectrum is then partitioned into critical bands according[3] to Table 3 1, and the energy in each critical band is summed

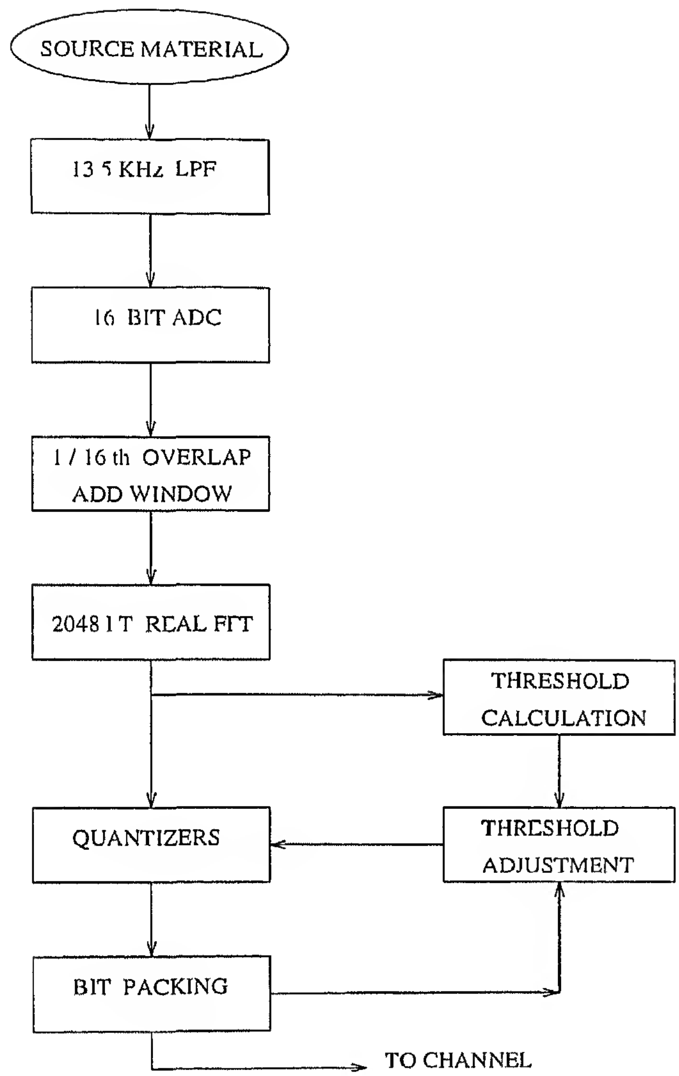


Figure 4 6 **Block Diagram Of Perceptual Transform Coder**

$$B_i = \sum_{\omega=bl}^{bh} P(\omega) \qquad (4\ 5)$$

where

bl = lower boundary of critical band i

bh_i = upper boundary of critical band i

B_i = energy in critical band i

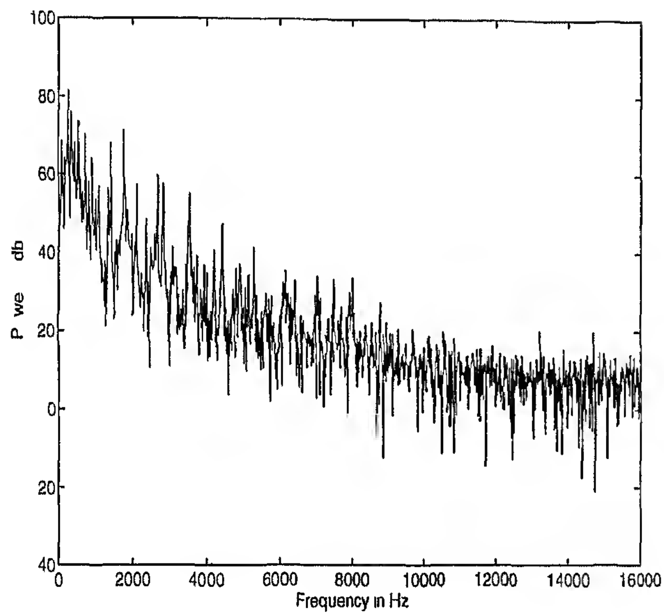


Figure 4 7 **Power Spectrum**

Fig 4 10 shows power spectrum and critical band spectrum *(Bark Spectrum)* for a block of input signal As discussed in section 3 1,critical band rate scale follows a linear scale up to about 500 Hz and then a logarithmic frequency scale above 500 Hz

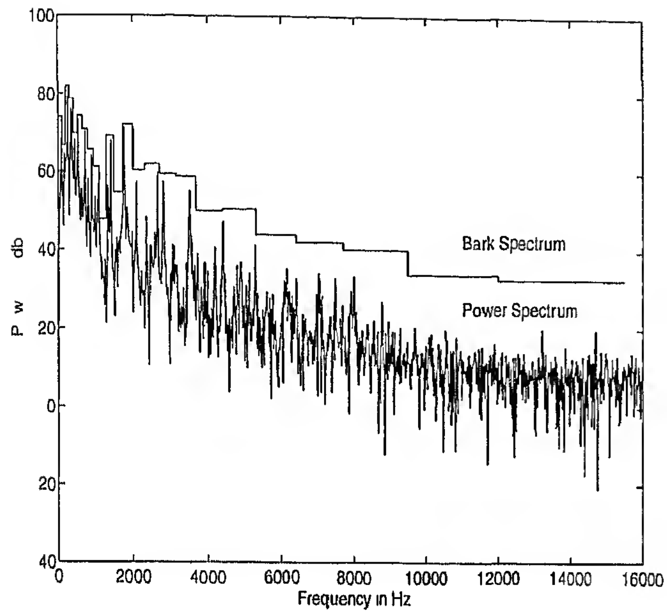


Figure 4 8 **Power and Bark Spectrum**

4 4 Spectral Flatness Measure

In order to determine the noise like or tone like nature of the signal, the *Spectral Flatness Measure(SFM)* is used The SFM is defined as the ratio of the geometric mean(Gm) of the power spectrum to the arithmetic mean(Am) of the power spectrum

$$SFM = \frac{lim_{N \to inf} (\prod_{k=1}^{N} P_k)^{\frac{1}{N}}}{lim_{N \to inf} \frac{1}{N} (\sum_{k=1}^{N} P_k)} \tag{4 6}$$

This ratio, by definition is not greater than unity SFM in decibels is given by

$$SFM_{dB} = 10 \log_{10} \frac{Gm}{Am} \tag{4 7}$$

SFM is then used to generate a coefficient of tonality α, to determine whether the signal is entirely tonelike or noise like α is given by

$$\alpha = min\left(\frac{SFM_{dB}}{SFM_{dBmax}}, 1\right) \tag{4.8}$$

With $SFM_{dBmax} = -60dB$, SFM obtained with three successive frames of 16 bit PCM monophonic audio signal of 15 KHz bandwidth sampled at 32 KHz are as tabulated as follows

Frame No	*SFM dB*	α	*Tonal /nontonal*
1	35.577	0.59	Tonal
2	23.560	0.39	Tonal
3	32.021	0.53	Tonal

Table 4.1 **Tonality Measure**

The complete code that is written using *Matlab* Package on Pentium PC @ 133 MHz for implementation of the Transform Coder and for computing the Spectral Flatness Measure and Bark Spectrum is given in Appendix

Chapter 5

Future Scope Of The Thesis

5.1 Scope Of The Thesis

The Transform Coders using perceptual noise criteria for efficient coding of stereophonic audio signals is underway. More recent work suggest that the Perceptual Transform Coder could be used at a lower bit rate of 3 bits per sample for transparent coding of signal sources. Examination of some 20 KHz bandwidth signals sampled at 44.1 KHz suggest that the same or slightly fewer bits per sample should provide transparent coding. The reason is that the lower energy at very high frequencies requires fewer than average bits for the additional bandwidth, resulting in the lower rate measured in bits per sample. Work related to increased coding efficiency, efficient stereo encoding and more effective compression/packing algorithms is underway.

5.2 Introduction To DAB

5.2.1 Why DAB System?

The use of digital transmission techniques for the terrestrial broadcasting of radio programs was previously prevented by the large bandwidth requirements that is, high frequency demands. As a result of modern audio compression techniques which exploit the psychoacoustic properties of the human ear, the present state of the art is capable of achieving

considerable reductions in the bit rate transmitted. Therefore an opportunity will arise for transmitting a digitally encoded stereo program in an FM channel bandwidth. The main objective of DAB(Digital Audio Broadcasting) is to achieve error free signal transmission with better audio quality and a minimum use of bandwidth. Audio quality must be an improvement over the basic characteristics(such as noise and distortion) of FM whose radio channels are highly impaired by multipath reception. Therefore DAB would be an efficient successor of the terrestrial FM broadcasts by occupying the VHF band of 87.5 to 108 MHz without losing[9] the capacity to broadcast the existing number of FM programs.

Radio channels using digital transmission techniques will permit the operation of single frequency networks in the future. These digital single frequency networks will permit substantial frequency economy. Broadcasting a radio program throughout a service area only requires one single transmission channel[2] or a part of a corresponding multiplex signal. The larger the service area, the less frequency volume is required for broadcasting a program. For this reason DAB will require only about one third of the frequency volume to broadcast programs, compared with FM today. In order to reuse frequency blocks as often as possible and due to other physical conditions relating to digital single frequency networks transmitter power compared with FM can be reduced considerably. Only low transmitter power permits the reuse of the same frequency block at distances that are not too great. Moreover directional antennas with levels of up to 20 dB must be used at the edges of service areas to minimize reduced radiation patterns.

5.2.2 Merits Of The DAB System

- The DAB single frequency networks will exploit the resources of wireless transmission with extraordinary efficiency. This applies in particular to programs which are broadcast nationwide or for large service areas.

- DAB will also provide listeners in the long term with a much greater program variety for terrestrial reception with mobile and portable radios. Full coverage broadcasting of programs that are error free and of excellent quality will be ensured at the start of DAB. That is, in future, DAB will become the *CD RADIO*.

Appendix

```
            / Transform Coder
 N=2048
 m=input( Enter the number of frames  ')
 fid1=fopen('m tim' 'r')
 dum = fread(fid1 20 'short')
 x = fread(fid1 2048*m 'short')
 x1=reshape(x N m)
                         / hanning window
 for c1=1 256
g(c1)=cos(2*pi*c1/257)
h(c1)=sqrt(0 5*(1-g(c1)))
 end
 t=ones(1 1792)
 u=t(1  )
 w(129 1920)=u
 for c2=1 128
       w(c2)=h(c2)
w(1920+c2)=h(128+c2)
 end
 c=1 2048
 for l=1 m
   W(c)=w
   F(1 2048)=W(c),
   K =x1(1 N l) *F',
   ft=fft(K N)
   reft=ft
   recp=real(ft)
   imcp=imag(ft)
   ft(1)=ft(1)+ft(1025)*i
   sp=ft(1 1024)
   ft(1025 2048)=[],
```

```
 re=real(sp)
 ım=ımag(sp)
                    / Spectral peak calculatıon & Bıt rate calculatıon
sr=abs(re)
sı=abs(ım)
sql=0 0
thȷ=7 02
count=1 8
for ınd=1 128
    srm=max(sr(count))
    sım=max(sı(count))
    p(ınd)=m1x(srm sım)
    pdb(ınd)=20*log10(p(ınd))
    pquan(ınd)=0 664*round((pdb(ınd)/0 664))
    pval(ınd)=10^(pquan(ınd)/20)
    ql(ınd)=2*round(pval(ınd)/thȷ)+1
    nbt(ınd)=round(log2(ql(ınd)))
    count=count+8
 end
  ȷȷ=1 16
 for ıı=1 128
     for ȷ1=ȷȷ
          rec(ȷ1)=nbt(ıı)
     end
 ȷȷ=ȷȷ+16
 end
 sql=sum(ql),
 totb=sum(rec),
 whıle totb > 14320
    thȷ=thȷ*1 005
    for ıd1=1 128,
        ql(ıd1)=0 0
        nbt(ıd1)=0 0
```

```
  end
  for id1=1 128
      ql(id1)=2*round(pval(id1)/thj)+1
      nbt(id1)=round(log2(ql(id1)))
  end
  j2=1 16
  for i1=1 128
      for j3=j2
          rec(j3)=nbt(i1)
      end
      j2=j2+16
   end
   sql=sum(ql)
   totb=sum(rec)
end
while totb < 14320
   thj=thj/1 002
   for id2=1 128,
       ql(id2)=0 0
       nbt(id2)=0 0
   end
   for id2=1 128,
       ql(id2)=2*round(pval(id2)/thj)+1
       nbt(id2)=round(log2(ql(id2)))
   end
   j4=1 16
   for i2=1 128
       for j5=j4
           rec(j5)=nbt(i2)
       end
       j4=j4+16,
   end
   sql=sum(ql)
```

```
    totb=sum(rec)
 end
lev=2  rec
i3=1 16
for i4=1 128
    for i5=i3
        nlev(i4)=max(rec(i3))
    end
    i3=i3+16
end
for i6=1 128
    step(i6)=(2*pval(i6))/2 (nlev(i6))
end
j6=1 16
for i10=1 128
    for j7=j6
        ssiz(j7)=step(i10)
    end
    j6=j6+16
end
retrs=recp',
imtrs=imcp'
for i7=1 2048
    requan(i7)=ssiz(i7)*round(retrs(i7)/ssiz(i7))
    imquan(i7)=ssiz(i7)*round(imtrs(i7)/ssiz(i7))
end
for i11=1 2048
    redat(i11)=(requan(i11))+(imquan(i11)*i),
end
redatt=conj(redat')
s=ifft(redatt,N)
e=real(s),
rw=e *F'
```

```
   fid2=fopen('tc dat' 'ı')
   count1=fprintf(fid2 '%5 4f\n  rw)
   rwt=rw'
            / 16 bit PCM  to 'b' bit PCM conversion
   b=input('Lnter the number of bits  ')
   step=(2*32767)/(2^b)
   for i12=1 2048
       bit8(i12)=step*round(x1(i12 1)/step)
   end
   bitt=bit8'
   fid3=fopen('coder dat' 'ı')
   count2=fprintf(fid3 '%5 4f\n' bitt)
   %pause
   c=c+1920
   %sound(x1 32000)
   /pause(5)
   /sound(rw 32000)
   fclose('all')
end
   figure(1)
   subplot(1 2,1) plot(x1(1 2048 1))
   xlabel('Sample Number') ylabel('Amplitude')
   subplot(1 2 2) plot(rw(1 2048)),
   xlabel('Sample Number'),ylabel('Amplitude'),
   figure(2),
   subplot(1,2 1) plot(x1(1 2048,1)),
   xlabel('Sample Number') ylabel('Amplitude')
   subplot(1 2 2) plot(bitt(1 2048))
   xlabel('Sample Number') ylabel('Amplitude')
   %print txcod -ffigure(1) -deps2
   %printf cod8 -ffigure(2) -deps2
```

% Perceptual Transform coder

```
N=2048
fs=32000
m=input('Enter the number of frames ')
fid1=fopen('m tim' 'r')
dum = fread(fid1 20 'short')
x = fread(fid1 2048*m 'short')
x1=reshape(x N m)
                    % hanning window
for c1=1 256
g(c1)=cos(2*pi*c1/257)
h(c1)=sqrt(0 5*(1 g(c1)))
end
t=ones(1 1792)
u=t(1 )
w(129 1920)=u
for c2=1 128,
     w(c2)=h(c2)
w(1920+c2)=h(128+c2)
end
c=1 2048,
  % fft & psd calculation
for l=1 m
  W(c)=w,
  F(1 2048)=W(c)
  K =x1(1 N,l) *F'
  ft=fft(K,N),
  ft(1)=ft(1)+ft(1025)*i,
  sp=ft(1 1024)
  ft(1025 2048)=[],
  freq=16000*(0 1023)/1024,
```

```
fqt=freq'
re=real(sp)
im=imag(sp)
%pw=(re).^2 + (im).^2
pw=(sp.*conj(sp))/length(sp)
pwdb=10*log10(pw)
fid2=fopen('pw.dat','w')
fprintf(fid2,'%5.3f \t\t%6.3f\n',fqt,pw)
figure(1)
plot(freq,pwdb)
xlabel('Frequency in Hz'),ylabel('Power in db')

        % calculation of CB energy

fid3=fopen('lim.dat','r')
lt=fscanf(fid3,'%4d')
c3=1:2
cnt=lt(c3)
B=zeros(1,24)
slp=0.0
fid4=fopen('fqlim.dat','r')
llt=fscanf(fid4,'%5d')
cc3=1:2,
ccnt=llt(cc3)

    for c4=1:24,
        for c5=cnt(1):cnt(2),
            slp=slp+pw(c5)
        end
        for cc5=ccnt(1):ccnt(2),
            bk(cc5)=slp
        end
        cc3=cc3+1
```

```
        ccnt 11t(cc3)
        B(c4)  1p
         1p 0 0
        c3-c3+1
        cnt  lt(c3)
        cnt(1) cnt(1)+1
        c4-c4+1
    end
    bdb-10*log10(bk)
    figure(?)
    plot(freq pwdb 'y ' 1 15500 bdb 'r-')
    gtext('Power Spectrum') gtext('Bark Spectrum')
    xlabel('Frequency in H ') ylabel('Power in db')


                      % calculation of SFM
  div=1/length(pw)
  gm=0 0
  for cc8=1 1024
      gm=gm+10*log10(pw(cc8))
  end
  Gm=div*gm,                        % Geometric mean
  Am=10*log10(mean(pw))             % Arithmetric mean
  SFM=Gm-Am
  alfa=min((SFM/( 60)) 1)           % tonality index

 fclose('all')
 c=c+1920
end
 figure(3)
 plot(1 2048,w 'y ' 1921 3968,w 'y-',3841 5888,w 'y-')
 xlabel('Sample number'),ylabel('Amplitude'),
 %print psp -ffigure(1) -deps2,
```

```
/print bsp -ffigure(2) -deps2
/print win -ffigure(3) -deps2

          / SNR Calculation

fid1=fopen('m tim' 'r')
dum = fread(fid1 20  short')
x = fread(fid1 8000 'short')
fid2=fopen('tc dat' 'r')
tout=fscanf(fid2 '/f')
fid3=fopen('coder dat' 'r )
codout=fscanf(fid3 '%f')
ltc=length(tout)
inps=x(200 ltc) ^2
avin=(sum(inps))/(ltc-200)
qetc=x(200 ltc)-tout(200 ltc)
sqtc=qetc ^2
avtc=(sum(sqtc))/(ltc-199)
snrval=avin/avtc
snrtc=10*log10(snrval)
qecod=x(200 ltc)-codout(200 ltc)
sqcod=qecod ^2
avcod=(sum(sqcod))/(ltc-199)
snrcod=avin/avcod
snr8=10*log10(snrcod)
```

REFERENCES

1 **Eberhard Zwicker and U Tilmann Zwicker** Audio Engineering and Psychoacoustics Matching signals to the final receiver the human auditory system *J Audio Engg Soc* Vol 39 No 3 March 1991 pp 115 126

2 **Frank Muller Romer** 'Directions in Digital Audio Broadcasting *J Audio Engg Soc* Vol 41 No 3 March 1993 pp 158 173

3 **James D Johnston** Transform coding of audio signals using perceptual noise criteria *IEEE Journal on selected areas in communications* Vol 6 No 2 Feb 1988 pp 314 323

4 **N S Jayant and Peter Noll** *Digital coding of waveforms Principles and applications to speech and video* Englewood cliffs Prentice Hall Inc

5 **John G Beerends and Jan A Stemerdink** 'A perceptual audio quality measure based on a psychoacoustic sound representation *J Audio Engg Soc* Vol 40 No 12 Dec 1992 pp 963 978

6 **Karlheinz Brandenburg and Gerhard Stool** ISO MPEG 1 Audio A Generic standard for coding of high quality digital audio *J Audio Engg Soc* Vol 42 No 10 Oct 1994 pp 780 791

7 **M R Schroeder, B S Atal and J L Hall,** Optimizing digital speech coders by exploiting masking properties of the human ear' *J Audio Engg Soc Am* 66(6) Dec 1979, pp 1647 1652

8 **Seymour Shlien,** 'Guide to MPEG 1 audio standard' *IEEE Transactions on Broadcasting*,Vol 40 No 4 Dec 1994 pp 206 218

9 **Theodor A Prosch** A possible frequency planning method and related model calculations for the sharing of the VHF Band II (87 5 108 MHz) between FM and

DAB systems *IEEE Transactions on Broadcasting* Vol 37 No 2 Dec 1991 pp 55 63

10 **E Zwicker H Fastl** *Psychoacoustics Facts and Models*